Rapid deployment of a Machine Learning-based derived biomarker using publicly available data sources for covariate adjusted descriptive modeling.

Albert Taylor, Danielle Beaulieu, Dustin Pierce, Andrew Conklin, Jonavelle Cuerdo, Mike Keymer, David L. Ennist

Origent Data Sciences, Inc., Vienna, VA 22182

INTRODUCTION

Defining baseline characteristics for covariate-adjusted analyses to increase study power is not new. Multifactorial heterogeneous diseases including Amyotrophic Lateral Sclerosis (ALS), Alzheimer's Disease (AD), Parkinson's Disease (PD), and Huntington's Disease (HD) present a challenge in defining baseline covariates that add substantial benefit to study power. We developed a methodology for training machine-learning (ML) models that utilizes historical clinical trial patient data to provide a single prediction value to be used as a covariate in a trial's statistical analysis. We have adapted this methodology across disease areas and have developed a rigorous audit methodology based on best practices in the biostatistics field so that these new methods can be more easily shared across a field where rigorous vetting of new technologies is critical to adoption.

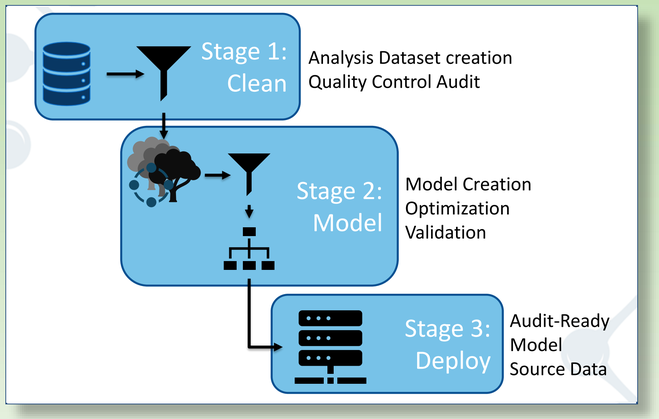
OBJECTIVE

To demonstrate through clinical trial simulation:

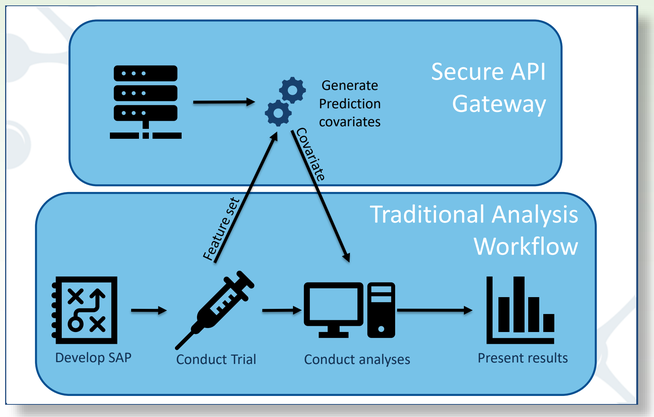
- A methodology for adopting rigorous methods for analysis dataset preparation for ML modeling
- A practical application of ML models to traditional biostatistical analysis
- A scalable approach that is applicable to multiple heterogeneous disease areas in which a suitable covariate is lacking

APPROACH

Standardized Workflow for Developing Machine Learning Models from Disease-Specific Clinical Trial Databases



Clinical Trial Workflow with ML Covariate Adjustment

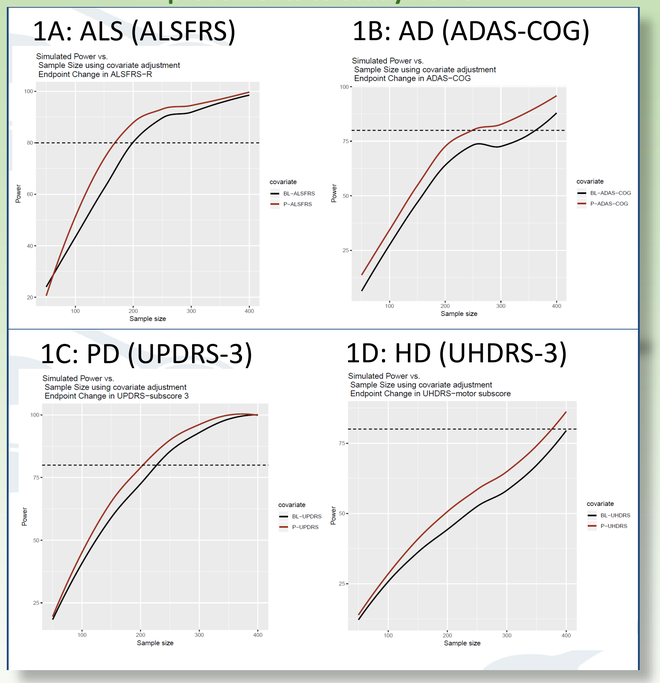


CONTACT

Albert Taylor, Director of Research
ataylor@origent.com

SIMULATION RESULTS

Clinical Trial Power Simulations Reveal Consistent Improvements to Study Power



Improvement to Study Power and reduction in variance (N = 150).

	ALS	AD	PD	HD
Power Boost	27.6%	26.2%	14.8%	20.6%
MSE Reduction	-18.1%	-11.7%	-3.6%	-10.2%

ACKNOWLEDGEMENTS

We thank the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) database, the Coalition Against Major Diseases (CAMD) database, and the Parkinson's Progression Markers Initiative (PPMI) database for access to data.